Open Access to Publications and to Data from Publicly Funded Research

Ireland and the World

Alan Smeaton
Professor of Computing and Founding Director of the Insight Centre for Data Analytics at Dublin City University

Open Access is a movement within scholarly research across all disciplines whereby the outputs, including publications and data, are made freely available to everyone without restriction. In the case of Open Access to data, it also covers data from public bodies like local governments. In Ireland we are slowly making progress with open access to publications and poor progress to research data while our Open Access to data from public bodies is progressing well.

Open access is a term normally reserved for describing how the outputs of research – covering publications and data – can be made openly accessible without restriction, to anybody. Traditionally, research outputs have typically been disseminated by publishing books or journal papers using the support and services of (academic) publishers, which have grown into huge businesses on the back of that. Conference or workshop presentations are also published, again by traditional publishers, and around this has grown the business model by which publishers operate: selling journal subscriptions and books to libraries, and charging a lot for doing so.

Open access is a noble aspiration which goes against this well-established model and has been enabled by the emergence of the internet and digital technologies, especially over the last couple of decades. Open access ensures that anyone, anywhere, at any time, can access research papers, book chapters, workshop papers, and even data used in experiments or analyses, and they can reuse this data or these publications without restriction.

In Ireland, almost all our higher education institutions (HEIs) have created online digital repositories or libraries for the research outputs of our scholars and scientists where their work has been funded by the taxpayer. These repositories are managed by our HEI libraries and have thereby created a new role and responsibility for the libraries. As a platform for disseminating research and scholarly outputs, institutional repositories are a safe bet in that they are funded and operate reliably; materials deposited in them are manually selected and curated by our librarians to ensure they do not break copyright or other restrictions when they are made available under open access. And because they are online digital repositories, the content is easily accessible.

The functionality offered by each HEI repository is basically the same. They accept 'deposits' from scholars and scientists, usually PDF files of papers, book chapters, or research theses. These are

manually checked by the library's curator for copyright conformance, then published in the online repository and thus made accessible. The underlying software systems are all interoperable and most are open source, so it doesn't matter if an institution uses DSpace, Greenstone, or EPrints (all of which are freely available), because these systems 'talk' to each other and allow content to be moved from one system to another.

Open Access is a noble aspiration which has been enabled by the emergence of the internet and digital technologies.

As an example of an institution's repository, Dublin City University's 'DORAS' currently has more than 7,000 items of content, each in PDF, broken into the following categories:

- Research article or paper (1,567)
- Book (18)
- Book section such as a chapter (214)
- Conference or workshop paper (2,462)
- Monograph (80)
- Research thesis (2,571)
- Working paper (128).

The research thesis category is an interesting one. It is there because DCU has, for many years, insisted that research theses from graduating students be made available on DORAS. Of the 2,571 theses, 1,830 are PhD theses, some going back to 1985, while the rest are at master's degree level.

One of the advantages of open source software is that statistics on access are automatically generated and can be made available to researchers for things like project reviews or funding applications. In the case of DORAS, for the calendar year 2016 alone, there were over 924,000 downloads of the PDFs of these 7,000 publications, which is part of a year-on-year increase in access.

In Ireland, almost all our Higher Education Institutions have created online digital repositories or libraries for the research outputs of our scholars and scientists where their work has been funded by the taxpayer.

Open access through deposit in institutional repositories is called 'green open access'. It is not the only avenue for open access publication. Many traditional publishers now offer an open access model where the author pays, usually a four-figure sum, to have their article hosted, disseminated, and promoted using the weight and impact of the publisher; this is called 'gold open access'. While it is valid as a form of open access dissemination, the 'author pays' part dissuades or prevents many researchers and scholars from using it.

Institutional open access repositories are not the only such repositories available: many open access hosting services have sprung up recently, such as ResearchGate and Academia.edu. These do not charge authors for publishing, and they offer great support for networking on their websites.

So why do scholars and researchers publish under open access? The trend is driven by several reasons. Increasing access to publications means they will be downloaded more and read by more people, which means they will then be cited more by other scholars and researchers; this is borne out repeatedly by organisations from the Wellcome Trust to the journal Nature. One study found that articles published in Nature Communications are viewed three times more than non-open-access content.

Open access also leads to more rapid dissemination, so the scientific and scholarly processes move faster and the wheels turn more quickly when newly published material reaches its target audience. It also leads to wider collaboration and more interdisciplinary follow-on work. A personal anecdote will illustrate this.

Open access leads to more rapid dissemination, wider collaboration and more interdisciplinary follow-on work.

One of my PhD students published his thesis on DORAS, and part of his work developed an algorithm for detecting periodicity in lifelogs or digital records of everyday activities. The thesis was found and read by a colleague, whom I'd never met, who works in chronobiology at the Royal College of Surgeons in Ireland (RCSI). She introduced us to her former colleagues at Glasgow University who are performing a chronobiologial analysis of accelerometer data from 100,000 people in the UK. Our team in DCU has now been brought into that work in Glasgow, and we have applied our periodicity-detection algorithm to their dataset. This is leading to further publications, all because the PhD thesis is available under open access in DCU.

Another reason open access is increasingly popular is that it is seen as a deliverable from publicly funded research. Many research funding bodies now mandate that research output from projects they fund be made available under open access. Projects funded under the EU's Horizon 2020 programme must make their outputs openly available, though it does not count ResearchGate or Academic.edu as open access publishing. Irish research funding agencies, including the Irish Research Council (IRC) and Science Foundation Ireland (SFI), also oblige their funded researchers to publish outputs under open access. While SFI does not specify which repository should be used, the IRC is more direct, stating that it 'should ideally be a local institutional repository to which the appropriate rights must be granted to replicate to other repositories'.

Content [of institutional repositories] is regularly crawled by web spiders and indexed by search services, including Google Scholar.

At a recent launch of newly funded research, one of our funding agencies pointed to the fact that approximately 40 per cent of publications from its funding were available under open access – a good number but not good enough yet. The Insight Centre at DCU has 436 research publications from the last four years available as open access, representing over 90 per cent of our research outputs in that period. Among the missing ones, some are embargoed until after their date of publication or because they are regarded as being in press. There is no reason we can't aim for 100 per cent of outputs available under open access.

One of the reasons why institutional repositories lead to increased visibility of research is that their content is regularly crawled by web spiders and indexed by search services, including Google Scholar. This is done precisely because their content is trusted, curated, and will not cause problems with copyright. The repositories also structure their metadata in a way that is easy for web spiders to assimilate. All of this helps to square the circle, because content from these repositories is then more easily found by people searching the internet.

However, all is not perfect in the world of open access publication. There are dangers from what are known as 'predatory publishers'. These are businesses which have emerged recently with a pay-to-publish business

model for scholarly publication but whose quality of publication is very poor, with low editorial standards, sloppy presentation, and even in some cases fake editors and fake editorial boards. As with everything on the internet, authors and readers have to be careful about the websites they trust. There are literally hundreds of these open access predatory publishers, cashing in on the huge push among researchers and scholars to publish their output.

So far in this article I have discussed open access to publications but not to data. The push for open access to data comes from the requirement to make new research outputs easily replicable by others, to verify and check research results, and to extend and build on that research. Some academic publishers now require that any data used in experiments in research papers must be made available under open access. This is helping to change people's approach to open access.

Almost 5,500 datasets are now available from public bodies on Ireland's Open Data Portal... this is a fabulous resource.

Open access to data can refer to any kind of data, from scientific experiments to surveys to data from public bodies, including governments. In March 2014, Ireland hosted the third plenary meeting of the Research Data Alliance (RDA), a global body whose remit is to develop and promote the standards to allow research data and public data to be easily shared and exchanged, and to make such data easily discoverable. More than 500 delegates from across the world gathered in Croke Park to present findings and to discuss and advance this agenda.

In Ireland, Open Data Fingal was the first local government portal created to publish open data (which is the term used to describe open access to data from public sources). That led to the establishment of Dublinked, which published public data about the city of Dublin, which in turn led to the establishment of a national open data portal, data.gov.ie.

Following the RDA meeting in 2014, and recognising the increasing importance of open data, in 2015 the government announced the formation of an Open Data Governance Board, chaired by Emer Coleman. In summer 2017 it published Ireland's Open Data Strategy 2017–2022, developed in conjunction with the Open Data Unit in the Department of Public Expenditure & Reform. The strategy involved a wide range of stakeholder consultation, including public consultation. It identifies seven strategic themes, including broadening the range of public bodies publishing open data; improving the quality, quantity, and range of datasets; continued encouragement of the use of open data; evaluating the impacts, benefits, and risks of the initiative; and more. It also has an implementation plan with aggressive timescales.

Almost 5,500 datasets are now available from public bodies on Ireland's Open Data Portal, and this is constantly increasing. Like open access to our research publications, this is a fabulous resource. One may not think of many use cases for data on sightings of hares in Ireland in 2006–2007, as published by the National Biodiversity Data Centre, or for data on life expectancy in Ireland by age and gender in 1994, 2004, and 2014, as published by Eurostat, but it is when datasets are combined and cross-correlated that their value increases.

So as we watch open access to research outputs develop and grow in importance, and we see the increasing availability of open data, what should we look out for? Mostly it's about two things. Firstly, we should encourage those who have not yet embraced open access to publications, research data, and public data, to do so for their own and everyone else's benefit. This includes encouraging our research funding agencies to more strictly enforce the open access principles in their policies. Eventually this will reach a tipping point where open access becomes the norm because people can see its benefits. We're not quite there yet. The second thing we should watch for is the opportunities that open access offers to everyone – not just researchers and journalists but all citizens. Open access is about everyone's access to data and content from public bodies, whether research funding bodies or local authorities, and that is something we should all have an interest in.

New Director of Irish Research Council

Peter Brown has been appointed Director of the Irish Research Council. From 2015 to 2017, he was Deputy Director of the Council, and Senior Manager with the Higher Education Authority from 2003 to 2014. Previously, he held roles with Dublin Institute of Technology and in the private sector.

Since May 2017, following the departure of former Council Director Dr Eucharia Meehan, Mr Brown has been Interim Director.

Brown's priority in his role as Director will be "to consolidate the Council's unique role within the Irish research eco-system".

"By investing in exceptional individuals working at the cutting edge of new knowledge, we future-proof Irish research and its place in a globally-connected world," he said.

"I am looking forward to leading the Council on this agenda over the coming years, working closely with Council members, a very committed staff team, the Higher Education Authority and Department of Education and Skills and the wider set of research stakeholders."